Identification of key genes and biological processes contributing to colitis associated dysplasia in ulcerative colitis

Di Zhang[1,2], Pengguang Yan[1,2], Taotao Han[1,2], Xiaoyun Cheng[1,2] and Jingnan Li[1,2]

[1] Department of Gastroenterology, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China
[2] Key Laboratory of Gut Microbiota Translational Medicine Research, Chinese Academy of Medical Sciences, Beijing, China

ABSTRACT

Background. Ulcerative colitis-associated colorectal cancer (UC-CRC) is a life-threatening complication of ulcerative colitis (UC). The mechanisms underlying UC-CRC remain to be elucidated. The purpose of this study was to explore the key genes and biological processes contributing to colitis-associated dysplasia (CAD) or carcinogenesis in UC via database mining, thus offering opportunities for early prediction and intervention of UC-CRC.

Methods. Microarray datasets (GSE47908 and GSE87466) were downloaded from Gene Expression Omnibus (GEO). Differentially expressed genes (DEGs) between groups of GSE47908 were identified using the "limma" R package. Weighted gene co-expression network analysis (WGCNA) based on DEGs between the CAD and control groups was conducted subsequently. Functional enrichment analysis was performed, and hub genes of selected modules were identified using the "clusterProfiler" R package. Single-gene gene set enrichment analysis (GSEA) was conducted to predict significant biological processes and pathways associated with the specified gene.

Results. Six functional modules were identified based on 4929 DEGs. Green and blue modules were selected because of their consistent correlation with UC and CAD, and the highest correlation coefficient with the progress of UC-associated carcinogenesis. Functional enrichment analysis revealed that genes of these two modules were significantly enriched in biological processes, including mitochondrial dysfunction, cell-cell junction, and immune responses. However, GSEA based on differential expression analysis between sporadic colorectal cancer (CRC) and normal controls from The Cancer Genome Atlas (TCGA) indicated that mitochondrial dysfunction may not be the major carcinogenic mechanism underlying sporadic CRC. Thirteen hub genes (*SLC25A3*, *ACO2*, *AIFM1*, *ATP5A1*, *DLD*, *TFE3*, *UQCRC1*, *ADIPOR2*, *SLC35D1*, *TOR1AIP1*, *PRR5L*, *ATOX1*, and *DTX3*) were identified. Their expression trends were validated in UC patients of GSE87466, and their potential carcinogenic effects in UC were supported by their known functions and other relevant studies reported in the literature. Single-gene GSEA indicated that biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways related to angiogenesis and immune response were positively correlated with the upregulation of *TFE3*, whereas those related to mitochondrial function and energy metabolism were negatively correlated with the upregulation of *TFE3*.

Conclusions. Using WGCNA, this study found two gene modules that were significantly correlated with CAD, of which 13 hub genes were identified as the potential key genes. The critical biological processes in which the genes of these two modules were significantly enriched include mitochondrial dysfunction, cell-cell junction, and immune responses. *TFE3*, a transcription factor related to mitochondrial function and cancers, may play a central role in UC-associated carcinogenesis.

INTRODUCTION

Ulcerative colitis (UC) is a subtype of inflammatory bowel disease (IBD), which is characterized by long-standing and recurring chronic inflammation of the colonic mucosa. UC-associated colorectal cancer (UC-CRC) is the most severe and life-threatening complication, especially in patients suffering from UC for a long duration. Epidemiological data have revealed that the risk of UC-CRC increases with disease duration (*Bernstein et al., 2001*; *Bopanna et al., 2017*). Although more attention has been given to the colonoscopic screening of dysplasia and cancer in UC patients (*Laine et al., 2015*; *Rubin et al., 2019*), early diagnosis is difficult because of the flat appearance and multifocal lesions. Therefore, there is an urgent need to explore the molecular biological mechanisms underlying UC-associated carcinogenesis and to find new strategies for early diagnosis, treatment, and prevention of UC-CRC.

Early genetic changes in precancerous lesions have been reported to contribute to the initiation of carcinogenesis. For example, *Ju et al. (2020)* revealed that 70% of miRNA alterations occur during the transition from normal to a preneoplastic stage of breast cancer. Similarly, *Zhang et al. (2020)* found that most of the differentially expressed genes (DEGs) identified in high-grade intraepithelial neoplasia (HGIN) and early gastric cancer (EGC) compared to their paired controls have already changed in low-grade intraepithelial neoplasia (LGIN) lesions. In addition, they identified 22 coDEGs (co-up DEGs and co-down DEGs), which are thought to play crucial roles in gastric tumorigenesis and progression, during the stages of LGIN, HGIN, EGC, and gastric adenocarcinoma. This phenomenon offers the possibility of early diagnosis and treatment of cancers in patients with UC. In fact, some abnormal molecular events have also been demonstrated in the inflamed colonic mucosa of UC before any apparent histological evidence of colitis-associated dysplasia (CAD) or cancer (*Itzkowitz, 2006*; *Itzkowitz & Yio, 2004*; *Tang et al., 2012*). However, comprehensive analysis based on gene expression profiles to reveal the genetic changes that occur in UC and contribute to carcinogenesis is still lacking.

In the present study, using database mining, we explored the key genes and biological processes contributing to CAD that were dysregulated in UC. We conducted weighted gene co-expression network analysis (WGCNA) based on DEGs between CAD and control

groups and found two gene modules correlated with the progression of UC-associated carcinogenesis. Functional enrichment analysis of the genes in these two modules revealed the associated crucial biological processes. Thirteen hub genes were identified as the potential key genes. Furthermore, we investigated the changes in their expression in the UC and CAD groups.

MATERIALS & METHODS

Data selection and processing

By searching the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database, the microarray dataset GSE47908 based on the GPL570 platform was selected for analysis because it embraces the transcriptional profiles of colonic mucosa from both UC and CAD patients (*Bjerrum et al., 2014*). Specifically, it included data from 45 patients with UC (20 with left-sided colitis, 19 with pancolitis, and six with UC-associated dysplasia) and 15 healthy controls. After obtaining the data, principal component analysis (PCA) was performed to check and visualize the grouped data. For validation, another microarray dataset GSE87466 comprising the gene expression profiles of colonic mucosa from 87 patients with UC and 21 healthy controls was selected to check the gene expression trends in UC (*Li et al., 2018*). For comparison, we downloaded the transcriptome and clinical data of 480 patients with sporadic colorectal cancer (CRC) and 41 healthy controls from The Cancer Genome Atlas (TCGA) database (date limit: February 23, 2021), using "TCGAbiolinks" R package version 2.16.4. The gene expression profiles were normalized using the voom function in "limma" R package (*Ritchie et al., 2015*).

Differential expression analysis

Differential expression analysis was performed using the "limma" R package (*Ritchie et al., 2015*). The sum of the mean value of absolute $\log_2$ fold change (FC) and the two standard deviations of absolute $\log_2$ FC was used as the cut-off value of $\log_2$ FC. Genes with absolute $\log_2$FC >$\log_2$ FC cut-off and adjusted p value <0.05, were considered as DEGs. The results are visualized as volcano plots.

WGCNA

The expression profiles of DEGs between CAD and control group were extracted to construct weighted gene co-expression network using R package "WGCNA" (*Langfelder & Horvath, 2008*; *Zhang & Horvath, 2005*). First, the candidate soft-thresholding powers (1 to 30) were used to calculate the scale independence and mean connectivity using the pickSoftThreshold function. The first candidate power, whose degree of independence was >0.8, was selected as the proper power. Then, a co-expression network was constructed and modules were identified using the blockwiseModules function, with the parameters mergeCutHeight set to 0.28 and minModuleSize set to 30. Each module was assigned a unique color. Finally, the Pearson correlation coefficient and corresponding p value between each module's eigengene and phenotype were calculated. In addition to the UC and CAD phenotypes, another phenotype "Progress" was also included to quantify the dynamic process of UC-associated carcinogenesis. The phenotype "Progress" was assigned

the values of 0,1, and 2, to represent the control, UC, and CAD groups, respectively; this reflected the progressive process from normal to UC and then to CAD. It means that if a module is significantly correlated to "Progress", then this module is considered to be correlated to the progressive process from normal to CAD.

Identification of hub genes

For a specified module gene, the module membership (MM) calculated by signedKME function represents its degree of importance in the module, and gene significance (GS) calculated by the cor function represents the degree of correlation with the phenotype. By calculating MM and GS, the module genes with high MM and GS can be defined as hub genes. In the present study, thresholds of MM >0.9 and GS >0.7, were chosen to screen for the hub genes of each module (*Zou et al., 2019*).

Function enrichment analysis

Functional enrichment analysis was performed using the "clusterProfiler" R package (*Yu et al., 2012*). Significantly enriched Gene Ontology-Biological processes (GO-BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were identified using enrichGO and enrichKEGG functions, respectively. Gene set enrichment analysis (GSEA) was performed based on the gene list sorted by log_2FC obtained from differential expression analysis using gseGO and gseKEGG functions (*Subramanian et al., 2005*). Single-gene GSEA was conducted based on the gene list sorted by Spearman correlation coefficient between every gene and the specified hub gene to predict the significant biological processes and pathways associated with the hub gene. A *p* value <0.05 was considered significant.

RESULTS

GEO data overview

A flow chart of the present study is shown in Fig. 1A. As for GSE47908 dataset, 15 controls, 20 patients with left-sided colitis, and six patients with UC-associated dysplasia were enrolled in the present study, representing the control, UC, and CAD groups, respectively. Clinical data including sex, age, disease duration, Mayo score, Mayo endoscopic score, smoking status, and daily medication were extracted and compared using the chi-squared test (*Bjerrum et al., 2014*). Results showed that there were no significant differences in sex, disease duration, smoking status, or daily medication between the groups. As for age, Mayo score, and Mayo endoscopic score, only median and interquartile ranges (IQR) were known, and the corresponding *p*-values could not be calculated (Table S1). After processing, four outliers were removed and a total of 37 samples, including 13 controls, 18 patients with UC, and six patients with CAD, were enrolled. As shown in the PCA plot (Fig. 1B), the gene transcriptional profiles of these three groups were clearly distinct from each other. The PCA plot of gene expression profiles of 87 patients with UC and 21 healthy controls in GSE87466 dataset shown in Fig. 1C, also indicates a significant distinction between these two groups.
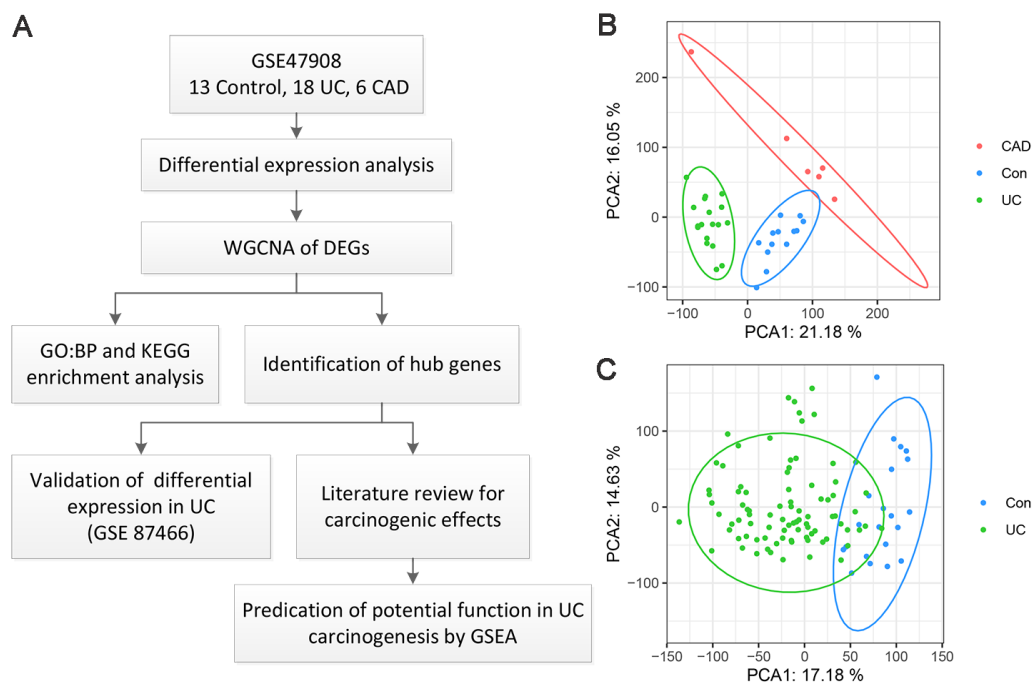
Figure 1 Overview of the work flow and GEO data. (A) Flow chart of the present study; (B) Principal component analysis plot of the gene expression profiles in GSE47908. Two principal components containing 37.23% of the variance illustrate separate clustering of the three group of samples; (C) Principal component analysis plot of the gene expression profiles in GSE87466. It shows the separate clustering of the two groups with a few samples overlapping by two principal component containing 31.81% of the variance.

Full-size 🖾 DOI: 10.7717/peerj.11321/fig-1

Differential expression analysis

To investigate the genetic changes during the progress from normal to UC and then to CAD, differential expression analyses between UC and control, CAD and UC, and CAD and control groups in GSE47908 were performed. The results are presented in Fig. 2. There were 679 upregulated and 288 downregulated genes in UC compared to those in the control group (Fig. 2A, $\log_2$ FC cut-off of 1.092), 260 upregulated and 616 downregulated genes in CAD compared to those in the UC group (Fig. 2B, $\log_2$ FC cut-off of 1.073), and 305 upregulated and 462 downregulated genes in CAD compared to those in the control group (Fig. 2C, $\log_2$ FC cut-off of 0.804). Venn diagrams were then used to further detail the genetic changes during this dynamic process. There were 67 co-up DEGs (Fig. 2D) and 46 co-down DEGs (Fig. 2E) in UC vs. controls and CAD vs. controls. In addition, the Venn diagram shows that among 260 upregulated and 616 downregulated genes in CAD vs. UC, only 61 upregulated (Fig. 2D) and 180 downregulated genes (Fig. 2E) changed significantly compared to the control group. This indicates that some DEGs dysregulated during the progress from normal to UC may continue to change in CAD and even contribute to the initiation of CAD, while others might be only responsible for the development of UC and may change back to the normal level if the inflammation is controlled.
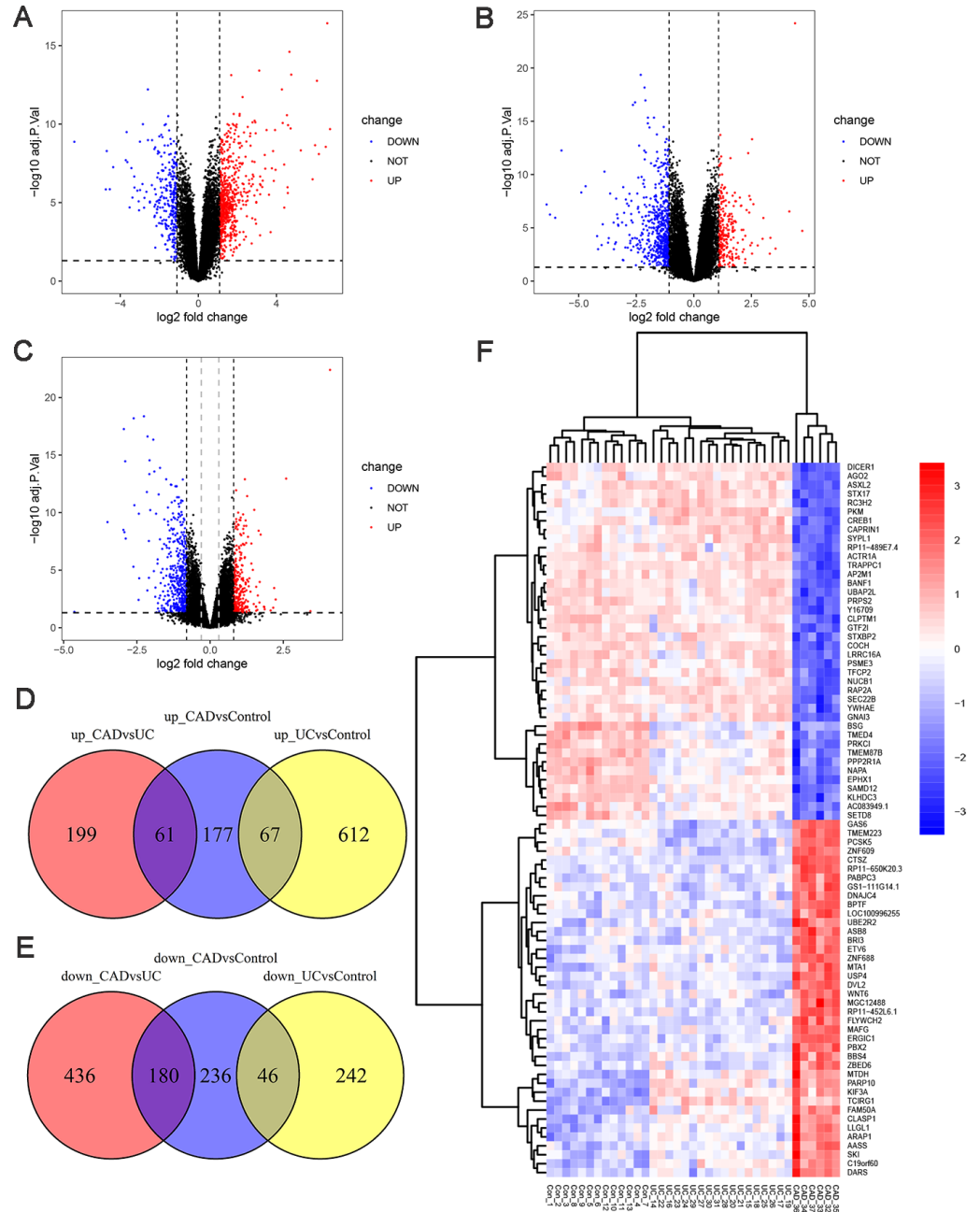
Figure 2 Differential expression analysis. (A) Volcano plot between UC and control groups. There are 679 up-regulated genes and 288 down-regulated genes in UC compared to controls when $\log_2$FC cut-off value was set to 1.092 by calculated. (B) Volcano plot between CAD and UC groups. There are 260 up-regulated genes and 616 down-regulated genes in CAD compared to UC when $\log_2$FC cut-off value was set to 1.073 by calculated. (C) Volcano plot between CAD and control groups. Black vertical line indicates $\log_2$FC of 0.804, and grey vertical indicates $\log_2$FC of 0.3. There are 305 up-regulated genes and 462 down-regulated genes in CAD compared to controls when $\log_2$FC cut-off value was set to 0.804, 2904 up-regulated genes and 2025 down-regulated genes when $\log_2$FC cut-off value was set to 0.3. (D) Venn diagram illustrating up-regulated genes in different groups. (E) Venn diagram illustrating down-regulated genes in different groups. (F) Heat map of top 40 up-regulated and 40 down-regulated DEGs between CAD and controls after sorted by adjusted p-value from smallest to largest.

Full-size 🖼 DOI: 10.7717/peerj.11321/fig-2

To gain further insights into the expression pattern of DEGs during the progress from normal to UC and then to CAD, the top 40 upregulated and top 40 downregulated DEGs between CAD and controls after sorting by adjusted *p*-value were visualized in a heat map (Fig. 2F). Row cluster analysis showed that these DEGs were grouped under two clusters: upregulated and downregulated genes in CAD compared to those in controls. However, when UC samples were taken into consideration, every cluster could be further divided into two sub-clusters, showing different changing trends of these DEGs during the progress from normal to CAD.

It should be noted that no co-up or co-down DEGs in UC vs. control groups and CAD vs. UC groups were identified in Venn diagrams, while the heatmap showed marked continuous changes in the expression patterns of some DEGs from control to UC and then to CAD. The traditional filter method for DEGs can potentially miss somekey information. To make our analysis more comprehensive, the cut-off criteria of $\log_2$ FC was set to 0.3, leading to the identification of 4929 DEGs between the CAD and control groups, which is a suitable number of DEGs for WGCNA analysis.

WGCNA

WGCNA was performed using 4929 DEGs between the CAD and control groups. By setting the soft threshold to 12 (Figs. 3A–3B), a total of six functional modules, apart from the gray module, were identified (Fig. 3C). The turquoise, red, and blue modules were positively correlated with CAD, and the brown, yellow, and green modules were negatively correlated with CAD. Furthermore, the green and blue modules were correlated with UC and CAD in the same direction respectively, and were found to be correlated with "Progress" with the highest correlation coefficients (green module: correlation coefficient = −0.75, *p*-value = 9e−08; blue module: correlation coefficient = 0.71, *p*-value = 9e−07) (Fig. 3D). This indicates that genes belonging to blue and green modules and the relevant biological processes begin to change at an earlier stage and continue to change throughout the whole process of progression from normal to UC and then to CAD.

Function enrichment analysis

The GO-BP enrichment analysis of each module is shown in Table S2. Green module genes were significantly enriched in biological processes related to mitochondrial function and energy metabolism (Fig. 4A). Consistent with this observation, KEGG pathway enrichment analysis indicated that the green module genes were significantly enriched in energy metabolism-related pathways, such as the TCA cycle, carbon metabolism, and oxidative phosphorylation (Fig. 4B). Blue module genes were significantly enriched in biological processes related to cell–cell junctions and immune responses, especially neutrophil-mediated immunity (Fig. 4C). The KEGG pathways that blue module genes were significantly enriched in included infection-related pathways, focal adhesion, tight junction, Hippo signaling pathway, Notch signaling pathway, and ErbB signaling pathway (Fig. 4D).

To further explain the role/s of mitochondrial dysfunction during the process of UC-associated carcinogenesis, GSEA was performed to analyze the alterations in
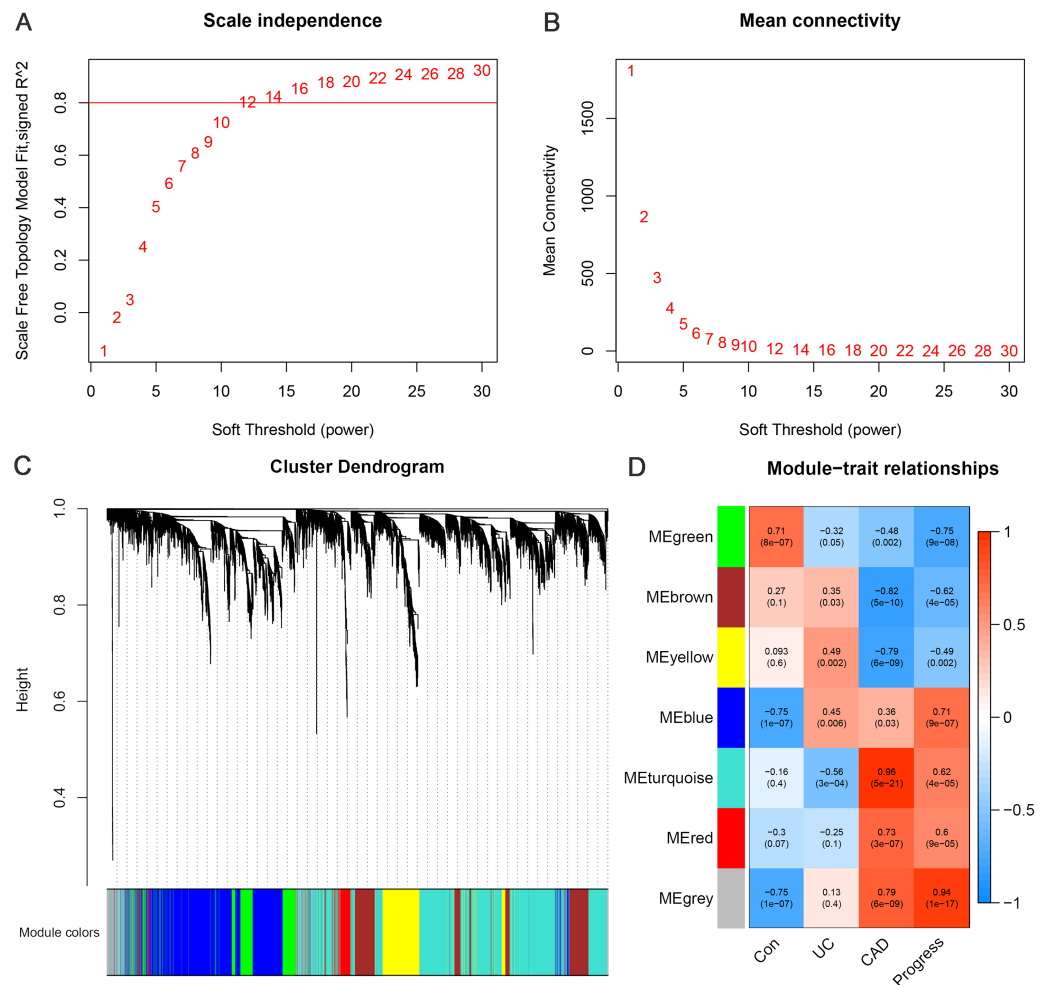
Figure 3 Weighted gene co-expression network analysis. (A) Network topology for different soft-thresholding powers. The scale-free topology can be attained at the soft-thresholding power of 12. (B) Cluster dendrogram and gene modules identified by WGCNA. (C) Correlation heat map of gene modules and traits, labelled with correlation coefficient and *p* value. The trait "Progress" means the progressive process from normal to UC, and then to CAD, which is assigned the value of 0, 1 and 2, to represent the control, UC and CAD group, respectively.

Full-size ⬛ DOI: 10.7717/peerj.11321/fig-3

mitochondrial function-related GO-BP in the UC and CAD groups. The results indicated that mitochondrial function-related GO-BP, including cellular respiration, oxidative phosphorylation, respiratory electron transport chain, and mitochondrial gene expression, were all significantly downregulated in UC (Fig. S1A) and CAD (Fig. S1B) groups as compared to those observed for the control group.

To explore the differences between UC-CRC and sporadic CRC in terms of the alteration of mitochondrial functions, the gene expression profiles of 480 sporadic CRC patients and 41 healthy controls from the TCGA database were analyzed. After differential expression analysis, GSEA of GO-BP was conducted based on the gene list sorted by $\log_2$ FC. The results showed that biological processes, including mitochondrial gene expression and
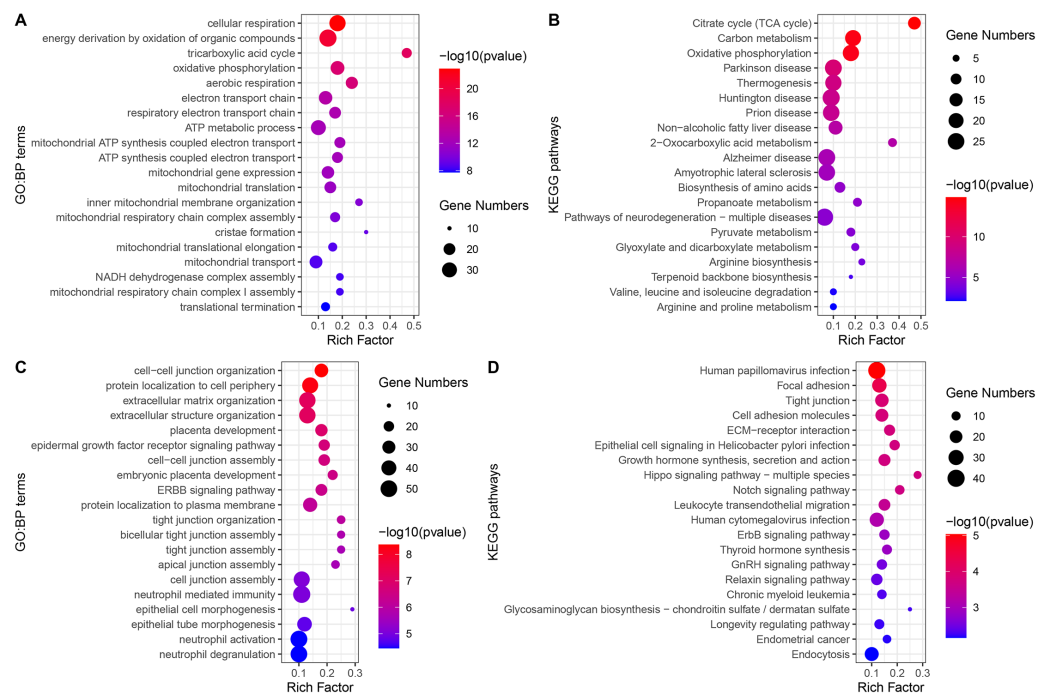
Figure 4 GO-BP and KEGG pathways enrichment analysis. (A) Biological processes enriched in green module genes. (B) KEGG pathways enriched in green module genes. (C) Biological processes enriched in blue module genes. (D) KEGG pathways enriched in blue module genes.

Full-size ☑ DOI: 10.7717/peerj.11321/fig-4

mitochondrial translation, were upregulated in sporadic CRC patients, and no other mitochondrial function-related biological processes were identified (Table S3). This indicates that the carcinogenic mechanism underlying sporadic CRC is different from that of UC-CRC.

Identification and validation of hub genes

Based on MM >0.9 and GS >0.7, seven genes (*SLC25A3*, *ACO2*, *AIFM1*, *ATP5A1*, *DLD*, *TFE3*, and *UQCRC1*) were identified as hub genes in the green module (Fig. 5A) and six hub genes (*ADIPOR2*, *SLC35D1*, *TOR1AIP1*, *PRR5L*, *ATOX1*, and *DTX3*) were identified in the blue module (Fig. 5C). The differential expression patterns of these 13 hub genes during the progression from normal to UC and then to CAD are displayed in Figs. 5B and 5D. All of them were observed to be continuously up/downregulated significantly, supporting their potential persistent roles in UC-associated carcinogenesis.

Considering the small sample size of the GSE47908 dataset, validation using another dataset was considered indispensable. To validate the changes in the inflammatory phase, differential expression analysis of the 13 hub genes was performed in UC patients from GSE87466 dataset (87 patients with UC and 21 controls). The results showed that the expression changes of hub genes in UC patients of GSE87466 dataset were consistent with those observed for patients belonging to GSE47908 dataset (Fig. 6). To further explore the potential association of hub genes with UC-associated carcinogenesis, a brief literature review was performed to search for the proof of their carcinogenic effects (*Bruggemann et*
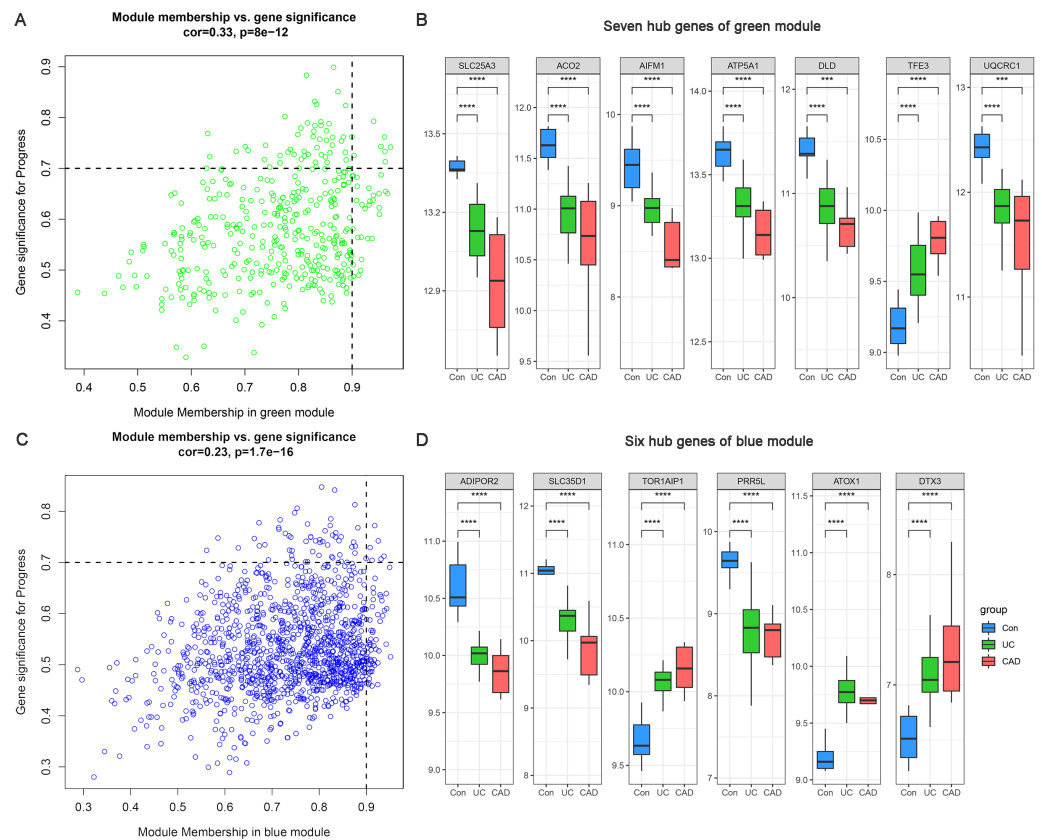
Figure 5 Hub genes of green and blue modules. (A) Identification of hub genes of green module by MM and GS filtering. (B) Differential expression patterns of seven hub genes of green module. (C) Identification of hub genes of blue module by MM and GS filtering. (D) Differential expression patterns of six hub genes of blue module.

Full-size 🖼 DOI: 10.7717/peerj.11321/fig-5

al., 2017; *Byeon et al., 2010*; *Chung et al., 2017*; *Ciccarone et al., 2020*; *Ding et al., 2020*; *Jana et al., 2020*; *Karginova et al., 2019*; *Khalil, 2007*; *Kim et al., 2019*; *Laiho et al., 2003*; *Li et al., 2019*; *Linehan et al., 2019*; *Liu et al., 2018*; *Millan & Huerta, 2009*; *Oehler et al., 2009*; *Perera et al., 2015*; *Rao et al., 2019*; *Seth et al., 2009*; *Thedieck et al., 2007*; *Viola et al., 2017*; *Wang et al., 2013*; *Wang et al., 2020*). The results are presented in Table 1. With the exception of *TOR1AIP1* and *PRR5L*, almost all other hub genes have been reported to be involved in cancers. Furthermore, five of the hub genes (*AIFM1*, *ATP5A1*, *UQCRC1*, *ADIPOR2*, and *ATOX1*) may play roles in the carcinogenesis of sporadic CRC. These known associations between hub genes and cancers suggest the possible carcinogenic effects of these hub genes in CAD.

Predication of the potential function of TFE3 in UC-associated carcinogenesis by GSEA

Among the proteins encoded by the hub genes, TFE3, a transcription factor, has been known to be involved in the onset and progress of cancers by regulating many biological processes, such as energy metabolism, lysosomal biogenesis, and immune response (*Beckmann,*
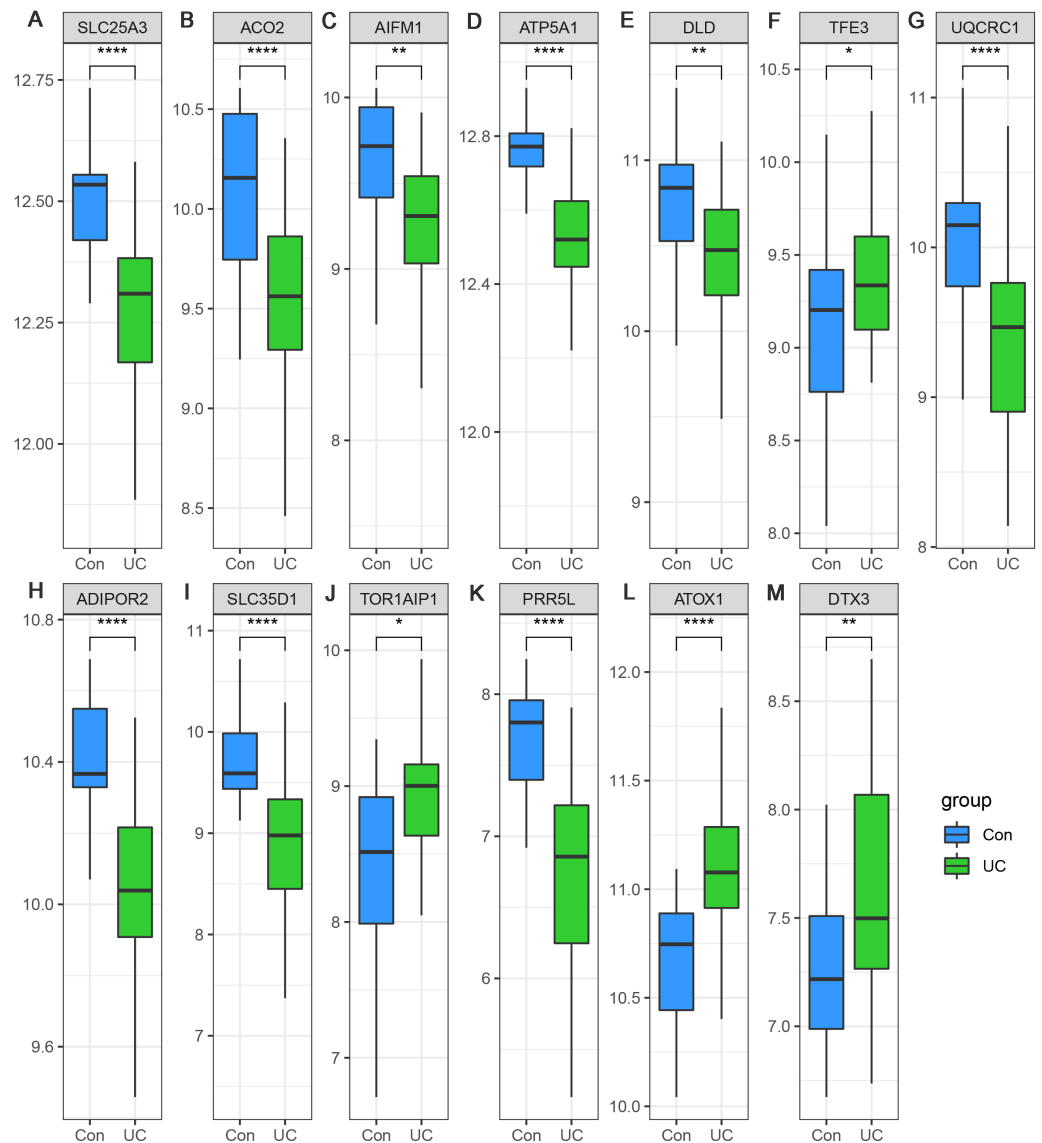
Figure 6 (A–M) Validation of the expression change of hub genes in UC from GSE87466.
Full-size ⊞ DOI: 10.7717/peerj.11321/fig-6

Su & Kadesch, 1990; *Perera et al., 2015*; *Willett et al., 2017*). In this study, we found that TFE3 was continuously upregulated during UC-associated carcinogenesis. However, the plausible roles of TFE3 in this process remain unclear. Single-gene GSEA revealed that biological processes and KEGG pathways related to angiogenesis and immune response were positively correlated with the upregulation of TFE3 (Figs. 7A, 7C), whereas those related to mitochondrial function and energy metabolism were negatively correlated with the TFE3 upregulation (Figs. 7B, 7D).

Table 1 Literature review of the proof for the carcinogenic effects of hub genes.

Genes	Encoding protein	Known function	Any proof for carcinogenesis	Any proof for CRC
SLC25A3	Phosphate carrier protein, mitochondrial	Transport of phosphate to mitochondrial matrix, involve in ATP synthesis	Yes (*Oehler et al., 2009*)	No
ACO2	Aconitate hydratase, mitochondrial	Catalyze the isomerization of citrate to isocitrate, involve in tricarboxylic acid cycle	Yes (*Ciccarone et al., 2020*; *Wang et al., 2013*)	No (*Laiho et al., 2003*)
AIFM1	Apoptosis-inducing factor 1, mitochondrial	Regulate apoptosis, and involve in mitochondrial respiratory	Yes (*Liu et al., 2018*; *Rao et al., 2019*)	Yes (*Millan & Huerta, 2009*)
ATP5A1	ATP synthase subunit alpha, mitochondrial	A component of the mitochondrial complex V in the respiratory chain, involve in ATP synthesis	Yes (*Bruggemann et al., 2017*)	Yes (*Seth et al., 2009*)
DLD	Dihydrolipoyl dehydrogenase, mitochondrial	Function as a dehydrogenase regulating energy metabolism, or as a protease	Yes (*Khalil, 2007*)	No
TFE3	Transcription factor binding to IGHM enhancer 3	Function as transcription factor, regulate many biological processes, such as energy metabolism, lysosomal biogenesis and immune response	Yes (*Linehan et al., 2019*; *Perera et al., 2015*)	No
UQCRC1	Cytochrome b-c1 complex subunit 1, mitochondrial	A component of the complex III in the respiratory chain, involve in ATP synthesis	Yes (*Wang et al., 2020*)	Yes (*Li et al., 2019*)
ADIPOR2	Adiponectin receptor protein 2	Regulates glucose and lipid metabolism, mediate increased AMPK and PPAR-alpha ligand activities	Yes (*Chung et al., 2017*)	Yes (*Byeon et al., 2010*)
SLC35D1	UDP-glucuronic acid/UDP-N-acetylgalactosamine transporter	Transport both UDP-glucuronic acid and UDP-N-acetylgalactosamine to ER lumen, participate in glucuronidation and/or chondroitin sulfate biosynthesis	Yes (*Viola et al., 2017*)	No
TOR1AIP1	Torsin-1A-interacting protein 1	Localize to the inner nuclear membrane, maintain the attachment of the nuclear membrane to nuclear lamina during cell division	No	No

Table 1 (*continued*)

Genes	Encoding protein	Known function	Any proof for carcinogenesis	Any proof for CRC
PRR5L	Proline-rich protein 5-like	Associate with mTORC2 complex that regulate cellular processes including survival and cytoskeleton organization	No (*Thedieck et al., 2007*)	No
ATOX1	Copper transport protein	Bind and deliver cytosolic copper to copper ATPase proteins, and be important in cellular antioxidant defense	Yes (*Karginova et al., 2019*; *Kim et al., 2019*)	Yes (*Jana et al., 2020*)
DTX3	Deltex E3 ubiquitin ligase 3	Function as an E3 ubiquitin ligase, regulate Notch signaling	Yes (*Ding et al., 2020*)	No
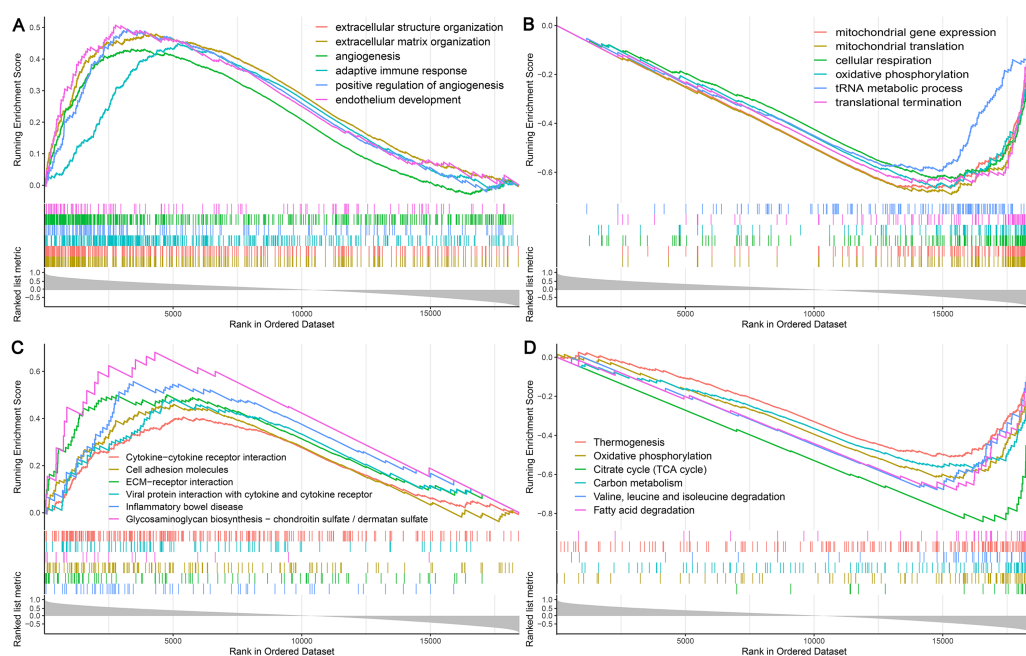
Figure 7 **Single-gene GSEA analysis of TFE3.** (A) Enriched biological processes of genes positively correlated with TFE3. (B) Enriched biological processes of genes negatively correlated with TFE3. (C) Enriched KEGG pathways positively correlated with TFE3. (D) Enriched KEGG pathways negatively correlated with TFE3.

Full-size ⬛ DOI: 10.7717/peerj.11321/fig-7

DISCUSSION

Most traditional research related to understanding of carcinogenic mechanisms focuses on comparison between tumor and normal tissue; however, this idea seems to be not fully applicable to UC-CRC considering its gradual molecular changes from inflammation to cancer. As UC-CRC arises due to the chronic inflammation of the colonic mucosa, it is logical to find the carcinogenic genes and biological processes associated with inflammation

during the progression of UC (*Itzkowitz & Yio, 2004*). In the present study, we identified two functional modules by WGCNA using the transcriptional profiles of colonic mucosa from healthy individuals and UC and CAD patients, which are correlated with UC and CAD in the same direction, suggesting their potential carcinogenic function when they contribute to inflammation. Thirteen hub genes (*SLC25A3*, *ACO2*, *AIFM1*, *ATP5A1*, *DLD*, *TFE3*, *UQCRC1*, *SLC35D1*, *TOR1AIP1*, *PRR5L*, *ATOX1*, and *DTX3*) were identified, and their continuous up/down-regulation supported their continuous roles during the process of progression from inflammation to carcinogenesis.

Functional enrichment analysis revealed that the green module genes were significantly enriched in biological processes and pathways related to mitochondrial function and energy metabolism. Consistently, almost all the hub genes of the green module (*SLC25A3*, *ACO2*, *ATP5A1*, *DLD*, and *UQCRC1*) were found to have involved in mitochondrial functions, such as the TCA cycle and ATP synthesis. Mitochondria have been discovered to play a multifunctional role in the pathogenesis of cancers (*Sajnani et al., 2017*). The well-known "Warburg effect" describes the shift in energy metabolism of cancer cells from mitochondrial respiration to glycolysis (*Warburg, 1956*). This shift may result from the mutation or dysregulation of genes related to energy metabolism, such as the genes encoding the enzymes involved in the TCA cycle. A decreased copy number of mtDNA has been found in lung cancer, liver cancer, gastric cancer, and CRC (*Lee et al., 2005*). That is to say, the downregulated hub genes related to mitochondrial respiration during UC inflammation may contribute to the transformation from UC to cancer by participating in the shift of energy metabolism. In addition, the alterations in these hub genes may lead to abnormal enzymatic reactions and oncometabolites, which in turn exert carcinogenic effects mainly through epigenetic regulation (*Nowicki & Gottlieb, 2015*). Another hub gene of the green module, *AIFM1*, is a well-known caspase-independent death effector released from mitochondria, and was later discovered to also play a role in oxidative phosphorylation (OXPHOS) by regulating complex I proteins post-transcriptionally. Both cell death and OXPHOS processes have been involved in cancer pathogenesis (*Liu et al., 2018*; *Rao et al., 2019*), but the role of *AIFM1* during UC-associated carcinogenesis needs further investigation.

TFE3 is a member of the MiT family of helix-loop-helix leucine zipper transcription factors that regulate their target genes by binding to E-box sequences in promoters. E-box sequences are mainly found in genes related to energy metabolism, including those involved in glycolysis and lipid metabolism. TFE3 has been reported to regulate mitochondrial dynamics and function in the liver (*Pastore et al., 2017*). Moreover, TFE3 has been validated as an oncogene in kidney and pancreatic cancers by virtue of its regulation of metabolic and non-metabolic pathways (*Linehan et al., 2019*; *Perera et al., 2015*). In the present study, single-gene GSEA revealed a link between TFE3 and mitochondrial function, while the latter was considered to be highly involved in UC-associated carcinogenesis. Therefore, it is logical to speculate that TFE3 may play a role in UC-associated carcinogenesis by regulating mitochondrial function. Further experimentation is required to confirm this hypothesis.

The genes encoding for proteins involved in cell–cell junctions and immune responses along with those involved in pathways, including the Hippo signaling pathway, Notch

signaling pathway, and ErbB signaling pathway, were significantly enriched in the blue module. Cell–cell junctions are vital for tissue homeostasis as they not only maintain the barrier function, but regulate complex cellular signaling networks related to cell proliferation and migration (*Garcia, Nelson & Chavez, 2018*). Disruption of tight junctions and subsequent immune dysfunction due to exposure to antigens play critical roles in UC pathogenesis. The Hippo pathway is an important signaling pathway regulating cell proliferation and apoptosis, and its dysregulation contributes to cancers. Moreover, the activity of the Hippo pathway is highly dependent on cell junctions (*Karaman & Halder, 2018*). Therefore, it is logical to think that the altered cell junctions and relevant signaling network during the inflammatory phase of UC may contribute to carcinogenesis (*Feigin & Muthuswamy, 2009*).

The obvious limitation of the present study is the small sample size; however, additional CAD or UC-CRC clinical specimens were not available for validation. The same expression trend of hub genes in another UC dataset and the reported relationships between these genes and cancers may help in demonstrating the reliability of this study. Further efforts are being made in this direction. This work is expected to provide new insights into the process of UC-associated carcinogenesis.

CONCLUSIONS

In conclusion, this study, using WGCNA, found two gene modules that were significantly correlated with the process of UC-associated carcinogenesis from inflammation to dysplasia. From these two modules, 13 hub genes (*SLC25A3*, *ACO2*, *AIFM1*, *ATP5A1*, *DLD*, *TFE3*, *UQCRC1*, *ADIPOR2*, *SLC35D1*, *TOR1AIP1*, *PRR5L*, *ATOX1*, and *DTX3*) were identified as key genes involved in UC-associated carcinogenesis. Functional enrichment analysis revealed the critical biological processes contributing to UC-associated carcinogenesis, mainly include mitochondrial dysfunction, cell–cell junction, and immune responses. TFE3, a transcription factor related to mitochondrial function and cancer, seem to play a central role in this process.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Author Contributions

- Di Zhang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Pengguang Yan performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Taotao Han analyzed the data, prepared figures and/or tables, and approved the final draft.
- Xiaoyun Cheng analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Jingnan Li conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The code is available in the Supplementary File. The data for analysis and validation are available at NCBI GEO: GSE47908 and GSE87466.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.11321#supplemental-information.

REFERENCES

Beckmann H, Su LK, Kadesch T. 1990. TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer muE3 motif. *Genes and Development* **4**:167–179 DOI 10.1101/gad.4.2.167.

Bernstein CN, Blanchard JF, Kliewer E, Wajda A. 2001. Cancer risk in patients with inflammatory bowel disease: a population-based study. *Cancer* **91**:854–862 DOI 10.1002/1097-0142(20010215)91:4<854::AID-CNCR1073>3.0.CO;2-Z.

Bjerrum JT, Nielsen OH, Riis LB, Pittet V, Mueller C, Rogler G, Olsen J. 2014. Transcriptional analysis of left-sided colitis, pancolitis, and ulcerative colitis-associated dysplasia. *Inflammatory Bowel Diseases* **20**:2340–2352 DOI 10.1097/MIB.0000000000000235.

Bopanna S, Ananthakrishnan AN, Kedia S, Yajnik V, Ahuja V. 2017. Risk of colorectal cancer in Asian patients with ulcerative colitis: a systematic review and meta-analysis. *The Lancet Gastroenterology and Hepatology* **2**:269–276 DOI 10.1016/S2468-1253(17)30004-3.

Bruggemann M, Gromes A, Poss M, Schmidt D, Klumper N, Tolkach Y, Dietrich D, Kristiansen G, Muller SC, Ellinger J. 2017. Systematic analysis of the expression of the mitochondrial ATP synthase (Complex V) subunits in clear cell renal cell carcinoma. *Translational Oncology* **10**:661–668 DOI 10.1016/j.tranon.2017.06.002.

Byeon JS, Jeong JY, Kim MJ, Lee SM, Nam WH, Myung SJ, Kim JG, Yang SK, Kim JH, Suh DJ. 2010. Adiponectin and adiponectin receptor in relation to colorectal cancer progression. *International Journal of Cancer* **127**:2758–2767 DOI 10.1002/ijc.25301.

Chung SJ, Nagaraju GP, Nagalingam A, Muniraj N, Kuppusamy P, Walker A, Woo J, Gyorffy B, Gabrielson E, Saxena NK, Sharma D. 2017. ADIPOQ/adiponectin induces cytotoxic autophagy in breast cancer cells through STK11/LKB1-mediated activation of the AMPK-ULK1 axis. *Autophagy* **13**:1386–1403 DOI 10.1080/15548627.2017.1332565.

Ciccarone F, DiLeo L, Lazzarino G, Maulucci G, Giacinto FDi, Tavazzi B, Ciriolo MR. 2020. Aconitase 2 inhibits the proliferation of MCF-7 cells promoting mitochondrial oxidative metabolism and ROS/FoxO1-mediated autophagic response. *British Journal of Cancer* **122**:182–193 DOI 10.1038/s41416-019-0641-0.

Ding XY, Hu HY, Huang KN, Wei RQ, Min J, Qi C, Tang H, Qin X. 2020. Ubiquitination of NOTCH2 by DTX3 suppresses the proliferation and migration of human esophageal carcinoma. *Cancer Science* **111**:489–501 DOI 10.1111/cas.14288.

Feigin ME, Muthuswamy SK. 2009. Polarity proteins regulate mammalian cell–cell junctions and cancer pathogenesis. *Current Opinion in Cell Biology* **21**:694–700 DOI 10.1016/j.ceb.2009.07.003.

Garcia MA, Nelson WJ, Chavez N. 2018. Cell-cell junctions organize structural and signaling networks. *Cold Spring Harbor Perspectives in Biology* **10**:a029181 DOI 10.1101/cshperspect.a029181.

Itzkowitz SH. 2006. Molecular biology of dysplasia and cancer in inflammatory bowel disease. *Gastroenterology Clinics of North America* **35**:553–571 DOI 10.1016/j.gtc.2006.07.002.

Itzkowitz SH, Yio X. 2004. Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. *American journal of Physiology. Gastrointestinal and liver physiology* **287**:G7–G17 DOI 10.1152/ajpgi.00079.2004.

Jana A, Das A, Krett NL, Guzman G, Thomas A, Mancinelli G, Bauer J, Ushio-Fukai M, Fukai T, Jung B. 2020. Nuclear translocation of Atox1 potentiates activin A-induced cell migration and colony formation in colon cancer. *PLOS ONE* **15**:e0227916 DOI 10.1371/journal.pone.0227916.

Ju Z, Bhardwaj A, Embury MD, Singh H, Gunaratne PH, Bedrosian I, Wang J. 2020. Integrative analyses of multilevel omics reveal preneoplastic breast to possess a molecular landscape that is globally shared with invasive Basal-like breast cancer (running title: molecular landscape of Basal-like breast cancer progression). *Cancers* **12**:722 DOI 10.3390/cancers12030722.

Karaman R, Halder G. 2018. Cell junctions in hippo signaling. *Cold Spring Harbor Perspectives in Biology* **10**:a028753 DOI 10.1101/cshperspect.a028753.

Karginova O, Weekley CM, Raoul A, Alsayed A, Wu T, Lee SS, He C, Olopade OI. 2019. Inhibition of copper transport induces apoptosis in triple-negative breast cancer cells and suppresses tumor angiogenesis. *Molecular Cancer Therapeutics* **18**:873–885 DOI 10.1158/1535-7163.MCT-18-0667.

Khalil AA. 2007. Biomarker discovery: a proteomic approach for brain cancer profiling. *Cancer Science* **98**:201–213 DOI 10.1111/j.1349-7006.2007.00374.x.

Kim YJ, Bond GJ, Tsang T, Posimo JM, Busino L, Brady DC. 2019. Copper chaperone ATOX1 is required for MAPK signaling and growth in BRAF mutation-positive melanoma. *Metallomics* **11**:1430–1440 DOI 10.1039/C9MT00042A.

Laiho P, Hienonen T, Mecklin JP, Jarvinen H, Karhu A, Launonen V, Aaltonen LA. 2003. Mutation and LOH analysis of ACO2 in colorectal cancer: no evidence of biallelic genetic inactivation. *Journal of Medical Genetics* **40**:e73 DOI 10.1136/jmg.40.5.e73.

Laine L, Kaltenbach T, Barkun A, McQuaid KR, Subramanian V, Soetikno R, Panel SGD. 2015. SCENIC international consensus statement on surveillance and management of dysplasia in inflammatory bowel disease. *Gastroenterology* **148**:639–651 DOI 10.1053/j.gastro.2015.01.031.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**:559 DOI 10.1186/1471-2105-9-559.

Lee HC, Yin PH, Lin JC, Wu CC, Chen CY, Wu CW, Chi CW, Tam TN, Wei YH. 2005. Mitochondrial genome instability and mtDNA depletion in human cancers. *Annals of the New York Academy of Sciences* **1042**:109–122 DOI 10.1196/annals.1338.011.

Li K, Strauss R, Ouahed J, Chan D, Telesco SE, Shouval DS, Canavan JB, Brodmerkel C, Snapper SB, Friedman JR. 2018. Molecular comparison of adult and pediatric ulcerative colitis indicates broad similarity of molecular pathways in disease tissue. *Journal of Pediatrics Gastroenterology and Nutrition* **67**:45–52 DOI 10.1097/MPG.0000000000001898.

Li W, Wubulikasimu G, Zhao X, Wang C, Liu R, Wang L, Zhu X, Chen Z. 2019. UQCRC1 downregulation is correlated with lymph node metastasis and poor prognosis in CRC. *European Journal of Surgical Oncology* **45**:1005–1010 DOI 10.1016/j.ejso.2019.02.025.

Linehan WM, Schmidt LS, Crooks DR, Wei D, Srinivasan R, Lang M, Ricketts CJ. 2019. The metabolic basis of kidney cancer. *Cancer Discovery* **9**:1006–1021 DOI 10.1158/2159-8290.CD-18-1354.

Liu D, Liu M, Wang W, Pang L, Wang Z, Yuan C, Liu K. 2018. Overexpression of apoptosis-inducing factor mitochondrion-associated 1 (AIFM1) induces apoptosis by promoting the transcription of caspase3 and DRAM in hepatoma cells. *Biochemical and Biophysical Research Communications* **498**:453–457 DOI 10.1016/j.bbrc.2018.02.203.

Millan A, Huerta S. 2009. Apoptosis-inducing factor and colon cancer. *Journal of Surgical Research* **151**:163–170 DOI 10.1016/j.jss.2007.05.020.

Nowicki S, Gottlieb E. 2015. Oncometabolites: tailoring our genes. *FEBS Journal* **282**:2796–2805 DOI 10.1111/febs.13295.

Oehler VG, Yeung KY, Choi YE, Bumgarner RE, Raftery AE, Radich JP. 2009. The derivation of diagnostic markers of chronic myeloid leukemia progression from microarray data. *Blood* **114**:3292–3298.

Pastore N, Vainshtein A, Klisch TJ, Armani A, Huynh T, Herz NJ, Polishchuk EV, Sandri M, Ballabio A. 2017. TFE3 regulates whole-body energy metabolism in cooperation with TFEB. *EMBO Molecular Medicine* **9**:605–621 DOI 10.15252/emmm.201607204.

Perera RM, Stoykova S, Nicolay BN, Ross KN, Fitamant J, Boukhali M, Lengrand J, Deshpande V, Selig MK. 2015. Transcriptional control of autophagy-lysosome function drives pancreatic cancer metabolism. *Nature* **524**:361–365 DOI 10.1038/nature14587.

Rao S, Mondragon L, Pranjic B, Hanada T, Stoll G, Kocher T, Zhang P, Jais A, Lercher A. 2019. AIF-regulated oxidative phosphorylation supports lung cancer development. *Cell Research* **29**:579–591 DOI 10.1038/s41422-019-0181-4.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**:e47 DOI 10.1093/nar/gkv007.

Rubin DT, Ananthakrishnan AN, Siegel CA, Sauer BG, Long MD. 2019. ACG clinical guideline: ulcerative colitis in adults. *American Journal of Gastroenterology* **114**:384–413 DOI 10.14309/ajg.0000000000000152.

Sajnani K, Islam F, Smith RA, Gopalan V, Lam AK. 2017. Genetic alterations in Krebs cycle and its impact on cancer pathogenesis. *Biochimie* **135**:164–172 DOI 10.1016/j.biochi.2017.02.008.

Seth R, Keeley J, Abu-Ali G, Crook S, Jackson D, Ilyas M. 2009. The putative tumour modifier gene ATP5A1 is not mutated in human colorectal cancer cell lines but expression levels correlate with TP53 mutations and chromosomal instability. *Journal of Clinical Pathology* **62**:598–603 DOI 10.1136/jcp.2009.064436.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**:15545–15550 DOI 10.1073/pnas.0506580102.

Tang A, Li N, Li X, Yang H, Wang W, Zhang L, Li G, Xiong W, Ma J, Shen S. 2012. Dynamic activation of the key pathways: linking colitis to colorectal cancer in a mouse model. *Carcinogenesis* **33**:1375–1383 DOI 10.1093/carcin/bgs183.

Thedieck K, Polak P, Kim ML, Molle KD, Cohen A, Jeno P, Arrieumerlou C, Hall MN. 2007. PRAS40 and PRR5-like protein are new mTOR interactors that regulate apoptosis. *PLOS ONE* **2**:e1217 DOI 10.1371/journal.pone.0001217.

Viola M, Bruggemann K, Karousou E, Caon I, Carava E, Vigetti D, Greve B, Stock C, DeLuca G. 2017. MDA-MB-231 breast cancer cell viability, motility and matrix adhesion are regulated by a complex interplay of heparan sulfate, chondroitin-/dermatan sulfate and hyaluronan biosynthesis. *Glycoconjugate Journal* **34**:411–420 DOI 10.1007/s10719-016-9735-6.

Wang P, Mai C, Wei YL, Zhao JJ, Hu YM, Zeng ZL, Yang J, Lu WH, Xu RH, Huang P. 2013. Decreased expression of the mitochondrial metabolic enzyme aconitase

(ACO2) is associated with poor prognosis in gastric cancer. *Medical Oncology* **30**:552 DOI 10.1007/s12032-013-0552-5.

Wang Q, Li M, Gan Y, Jiang S, Qiao J, Zhang W, Fan Y, Shen Y, Song Y. 2020. Mitochondrial protein UQCRC1 is oncogenic and a potential therapeutic target for pancreatic cancer. *Theranostics* **10**:2141–2157 DOI 10.7150/thno.38704.

Warburg O. 1956. On the origin of cancer cells. *Science* **123**:309–314 DOI 10.1126/science.123.3191.309.

Willett R, Martina JA, Zewe JP, Wills R, Hammond GRV, Puertollano R. 2017. TFEB regulates lysosomal positioning by modulating TMEM55B expression and JIP4 recruitment to lysosomes. *Nature Communications* **8**:1580 DOI 10.1038/s41467-017-01871-z.

Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**:284–287 DOI 10.1089/omi.2011.0118.

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* **4**: Article 17 DOI 10.2202/1544-6115.1128..

Zhang Y, Wu X, Zhang C, Wang J, Fei G, Di X, Lu X, Feng L, Cheng S, Yang A. 2020. Dissecting expression profiles of gastric precancerous lesions and early gastric cancer to explore crucial molecules in intestinal-type gastric cancer tumorigenesis. *Journal of Pathology* **251**:135–146 DOI 10.1002/path.5434.

Zou D, Li R, Huang X, Chen G, Liu Y, Meng Y, Wang Y, Wu Y, Mao Y. 2019. Identification of molecular correlations of RBM8A with autophagy in Alzheimer's disease. *Aging* **11**:11673–11685.